

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11231890 A**

(43) Date of publication of application: **27.08.99**

(51) Int. Cl.

G10L 3/00

G10L 3/00

G10L 3/00

G10L 5/02

G10L 5/04

(21) Application number: **10029516**

(71) Applicant: **OKI ELECTRIC IND CO LTD**

(22) Date of filing: **12.02.98**

(72) Inventor: **SUGIYAMA HIROBUMI**

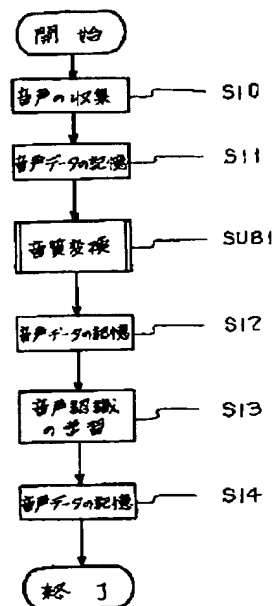
(54) **DICTIONARY PREPARATION METHOD FOR
SPEECH RECOGNITION**

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a dictionary preparation method for speech recognition, capable of preparing a speech recognition dictionary on the basis of useful speech data obtainable at low cost and in a short time.

SOLUTION: In this dictionary preparation method for speech recognition, speech data obtained through a sampling process at a step S10 is stored at a step S11, and speech data having new sound quality is generated via a subroutine SUB1, using original speech data so obtained. Furthermore, speech data after made to learn speech recognition at a step S13 on the basis of the generated speech data is stored at a step S14 for the preparation of a dictionary, thereby restraining the number of persons assignable to a sampling process and shortening a time required for the sampling process.

COPYRIGHT: (C)1999,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-231890

(43) 公開日 平成11年(1999) 8月27日

(51) Int. Cl. ⁶	識別記号	F I			
G10L 3/00	531	G10L 3/00	531	K	
	515		515	H	
5/02		5/02		B	
5/04		5/04		C	
				E	
審査請求 未請求 請求項の数 3 O L (全 7 頁)					

(21) 出願番号 特願平10-29516

(22) 出願日 平成10年(1998) 2月12日

(71) 出願人 000000295

沖電気工業株式会社

東京都港区虎ノ門1丁目7番12号

(72) 発明者 杉山 博文

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

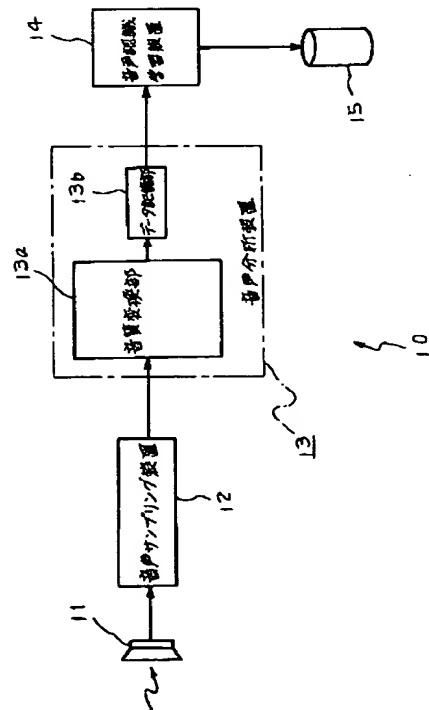
(74) 代理人 弁理士 香取 孝雄

(54) 【発明の名称】 音声認識の辞書作成方法

(57) 【要約】

【課題】 コストを抑えて従来よりも短時間で得られる有用な音声データを基に音声認識の辞書を作成する音声認識の辞書作成方法の提供。

【解決手段】 本発明の音声認識の辞書作成方法は、ステップS10 でサンプリングして得られた音声データをステップS11 ではこの音声データを保存し、得られたオリジナルな音声データを用いてサブルーチンSUB1で新たな音質を有する音声データを生成し、さらに、これらの音声データを基にステップS13 で音声認識を学習させた音声データをステップS14 で保存して辞書を作成して、サンプリングに要する人数を抑え、かつこれに要する時間の短縮も図っている。



音声認識辞書を作成するシステム構成の一例

【特許請求の範囲】

【請求項 1】 予め個々の話者の音声をサンプリングして得られた音声データを基に音声認識に用いる辞書の作成手順を示す音声認識の辞書作成方法であって、前記個々の話者の音声データを収録した後の前記音声データの品質を保ちながら、該音声データを用いて該音声データと異なる音質の音声データに変換させる音質変換工程と、
該音質変換工程により得られる新たな音声データと前記サンプリングして得られた音声データとを基に前記音声認識を学習する音声認識学習工程とを含み、
該音声認識学習工程によって得られた音声データを保存して音声認識の辞書にすることを特徴とする音声認識の辞書作成方法。

【請求項 2】 請求項 1 に記載の辞書作成方法において、前記音質変換工程は、前記音声データのスペクトルを分析するスペクトル分析工程と、
該スペクトル分析工程の結果を予め作成したベクトル量子化のコードと対応させ、該対応するベクトル量子化のコードを出力するベクトル量子化工程と、
該ベクトル量子化工程で得られたベクトル量子化のコードを任意に選んだ複数の話者間の音声の特徴の統計的な対応付けにより得られるスペクトル変換規則に基づいて前記スペクトル分析工程から得られたスペクトルを変換するスペクトル変換工程とを含むことを特徴とする音声認識の辞書作成方法。

【請求項 3】 請求項 2 に記載の辞書作成方法において、前記スペクトル変換規則は、複数の話者を入力話者とし、該入力話者の一人を音質変換対象話者に選び、残る各入力話者と前記音質変換対象話者の音声データからそれぞれ得られる音声のスペクトルの特徴を対応付けることを特徴とする音声認識の辞書作成方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、話者の音声認識に用いる辞書を作成する音声認識の辞書作成方法に関し、特に、不特定な話者の音声をサンプリングしてデータとして取り込み、この取り込んだデータを基に音声認識の学習による認識効果も期待できる辞書を作成する際に用いて好適なものである。

【0002】

【従来の技術】従来、不特定の話者の音声を認識する場合、最初に予め音声データを収録し、この収録に伴ってこれら多数の音声データが記憶装置に保存される。次に保存された音声データを基にたとえば、ワークステーションあるいはパーソナルコンピュータ等の情報処理装置を用いて音声認識が容易となるように経験を踏まえて新たな知識を習得するように音声認識の学習が行われる。この手順に従い学習した結果の音声データを保存して音声認識用の辞書が作成されている。

【0003】

【発明が解決しようとする課題】ところで、音声データの収録には、少なくとも数百人程度の人間、すなわち音源となる話者、あるいはエキストラが必要である。このため、これらの人員に支払う人件費がかなりかかることになる。この他、同様の辞書作成の費用面において、この音声データを正確に収録するには専用の機器が必要であるがこの専用の機器は高価なので、安価に済ませるためにこれらの機器はレンタル契約で借り受けて使用する場合がある。

【0004】また、この音声データの収録作業のコストダウンを鑑みた結果、この作業を他の部門や外部の会社等に発注すると、発注一納期の期間が短い場合、この納期が収録する音声データ数を制限してしまう場合がある。

【0005】このように音声データの収録には、経済的な負担と時間的な制約が大きな問題となる。

【0006】本発明はこのような従来技術の欠点を解消し、コストを抑えて従来よりも短時間で得られる有用な音声データを基に音声認識を行って辞書を作成する音声認識の辞書作成方法を提供することを目的とする。

【0007】

【課題を解決するための手段】本発明は上述の課題を解決するために、予め個々の話者の音声をサンプリングして得られた音声データを基に音声認識に用いる辞書の作成手順を示す音声認識の辞書作成方法であって、個々の話者の音声データを収録した後、音声データの品質を保ちながら、この音声データを用いてこの音声データと異なる音質の音声データに変換させる音質変換工程と、この音質変換工程により得られる新たな音声データと前記サンプリングして得られた音声データとを基に音声認識を学習する音声認識学習工程とを含み、この音声認識学習工程によって得られた音声データを保存して音声認識の辞書にすることを特徴とする。

【0008】ここで、音質変換工程は、音声データのスペクトルを分析するスペクトル分析工程と、このスペクトル分析工程の結果を予め作成したベクトル量子化のコードと対応させ、この対応するベクトル量子化のコードを出力するベクトル量子化工程と、このベクトル量子化工程で得られたベクトル量子化のコードを任意に選んだ複数の話者間の音声の特徴の統計的な対応付けにより得られるスペクトル変換規則に基づいてスペクトル分析工程から得られたスペクトルを変換するスペクトル変換工程とを含むことが好ましい。これにより、音声データの品質を低下させることなく異なる音質の音声データを生成している。

【0009】また、スペクトル変換規則は、複数の話者を入力話者とし、この入力話者の内の一人を音質変換対象話者に選び、残る各入力話者と音質変換対象話者の音声データからそれぞれ得られる音声のスペクトルの特徴

10

20

30

40

50

を対応付けることが好ましい。これにより、各入力話者と音質変換対象話者の中間的な音声生成される。

【0010】本発明の音声認識の辞書作成方法は、サンプリングして得られた音声データを用いて音質変換工程で新たな音質を有する音声データを生成し、これらの音声データを基に音声認識学習工程で音声認識を学習させた音声データを保存して辞書を作成することにより、ある話者からサンプリングされた音声データを用い新たな話者に対応する音声データが生成されるので、サンプリングに要する人数を抑えることができ、これによる時間の短縮も図れる。

【0011】

【発明の実施の形態】次に添付図面を参照して本発明による音声認識の辞書作成方法の一実施例を詳細に説明する。

【0012】本発明の音声認識の辞書作成方法は、たとえば音声認識装置の音声認識辞書を作成する際に適用される。この音声認識辞書の作成には、図1に示すように、辞書作成システム10を用いる。辞書作成システム10は、マイクロフォン11、音声サンプリング装置12、音声分析装置13、音声認識学習装置14、および記憶装置（辞書）15を備える。各部を簡単に説明すると、マイクロフォン11は、話者の音声を集音し音を電気信号である音声信号に変換して音声サンプリング装置12に出力する。

【0013】音声サンプリング装置12は、供給されたアナログの音声信号をデジタルの音声データに変換する。変換された音声データは、図示しない記憶装置に格納する。音声サンプリング装置12は収集した音声データを音声分析装置13に供給する。音声分析装置13は音質変換部13a、およびデータ記憶部13bとを含んでいる。音質変換部13aは、供給される音声データの中から2つの音声データを選んで、これらの選択した音声データの中間的な音声データを生成する。この音声データの作成の手順については後段で詳述する。音声データの作成手順の中で最後に行われるスペクトル変換処理によって得られた音声データをデータ記憶部13bに記憶させる。データ記憶部13bには音声サンプリング装置12から供給されたそのまの音質の音声データも記憶されている。

【0014】音声分析装置13は、この装置に供給された音声データおよび分析により新たに得られた音声データを音声認識学習装置14に出力する。音声認識学習装置14は、予め学習用に話された音声データを基に音声認識が迅速に処理できるようにこれまでの経験を踏まえて学習する。この学習で得られた結果が、音声認識学習装置14から記憶装置15に供給される。この供給されたデータを記憶することにより記憶装置15には、音声認識に用いる辞書データが格納されたことになる。このように作成した辞書データを音声認識装置の辞書に適用する。

【0015】なお、マイクロフォン11と音声サンプリング装置12は、一体的に構成して話者の集まる場所に携帯

してサンプリングを行い、音声分析装置13および音声認識学習装置14等と別々に扱うような使い方も可能である。このように構成することにより、場所に限定されずに集音することができる。

【0016】次に音声認識辞書の作成手順について図2および図3を参照しながら説明する。まず、音声認識辞書の作成開始にあたり話者をマイクの前に配置して音声サンプリング装置11をセットしてステップS10に進む。

【0017】ステップS10では、話者の音声を収集（すなわち、サンプリング）する。この音声サンプリングでは、たとえば音声をマイクロフォン11等を介して音声サンプリング装置11でアナログの電気信号に変換し、さらにこのアナログの電気信号をデジタル変換して音声データにしている。ここで、音声認識装置の音声認識の精度および処理の高速化を図れるように、音声認識の辞書には学習処理が施されている。この辞書を学習させるために発話者には、学習用の同一の文言・単語等を与えて発話させるようにしている。

【0018】次にステップS11では、ステップS10で得られた音声データを音声サンプリング装置11の記憶装置（図示せず）に記憶させる。ステップS10、S11の処理だけを繰り返して発話者の音声データの収集・格納をさせてもよい。このステップS11の後にサブルーチンSUB1に進む。

【0019】サブルーチンSUB1では、音声サンプリング装置11から音声分析装置13に取り込んだ音声データを用いて新たな音質の音声データを生成する（音質変換工程）。この生成される音声データは、複数の入力話者の内の中から一人を音質を変換させる変換対象話者を選び、残る入力話者とこの変換対象話者の音声データからそれぞれ得られるスペクトルの特徴を対応付けによりこの変換対象話者と入力話者間の中間的な音声データとして得られる。このサブルーチンSUB1については後段でさらに詳述する。音声データを生成した後、ステップS12に進む。ステップS12では、サブルーチンSUB1により得られた音声データをたとえばデータ記憶部13bに格納する。この格納により音声データの量が大幅に増加する。

【0020】次にステップS13では、増加した音声データも含んで保存されているデータ記憶部13bから出力される音声データに基づいて音声認識学習装置14で音声認識の学習を行う。このステップでは音声データを認識対象の発声単位によって分類するだけでなく、個人差に対する対処の仕方によって特定話者を認識するため学習用音声を用いて声の特徴を学習する（すなわち、教師あり学習を行う）。これにより、個人差によるスペクトルの変動を正規化したり、認識システムを発声者の声質に合うように適応化させる。

【0021】個人差の正規化・適応化の方法には、声帯波スペクトルと声道長の効果の個人差に着目したり、音素標準パターンの個人差による変動に音素間の共通性が

あることから、一部の音素標準パターンを少数の学習音声で学習し、かつ残りの音素は変換公式に応じて話者に適応化させる等の方法がある。この他、認識を行いながらその結果を用いて自動的に学習する、教師なし学習による音声認識方法等もある。この中で選んだ方法で音声認識の学習を行ってステップS14に進む。

【0022】ステップS14では、ステップS13での音声認識の学習処理結果を記憶する。したがって、音声認識学習装置14は記憶装置15に処理結果を供給する。この学習処理結果の記憶により、記憶装置15ないに音声認識用の辞書が完成する。この辞書の完成後、音声認識の辞書作成処理を終了する。

【0023】次にこの音声認識の辞書作成処理の特徴であるサブルーチンSUB1の音質変換処理について図3のフローチャートを参照しながら説明する。この音質変換処理は、特開平7-104792号公報に記載されている方法を用いている。この方法は、たとえば入力した個々の音声データを入力話者として扱う。そして、これら入力話者の内の一つの音声データを変換対象話者とした際にこれら入力話者と変換対象話者の中間に位置する音質の音声データを新たに生成する。この音声データを生成するため、まず、サブステップSS10に進む。

【0024】サブステップSS10では、入力話者のコードブックを生成する。コードブックとは、音声信号を効率的に表現するために用いられるパラメータ（音声特徴量）を算出し、この算出されたパラメータを統計的に分類した際に得られる分類表である。本実施例では分類表の作成に、線形予測分析・合成方式（LPC: Linear Predictive Codingの略で、以下、LPCという）を用いる。学習用に話された音声データに対してLPC分析をし、得られたLPCパラメータをクラスタリング（分類）を行う。クラスタリングには、たとえばLBG（Linde-Buzo-Gray）アルゴリズムを用いる。この処理の後、サブステップSS11に進む。

【0025】サブステップSS11では、入力話者のコードベクトル系列を生成する。サブステップSS10で生成した入力話者のコードブックに基づいて入力話者学習用音声データにLPC分析およびベクトル量子化を施す。ベクトル量子化とは、音声データをLPC分析して得られたLPCパラメータに最も似たスペクトル特徴量をコードブックから出力することである。これらの集まりをコードベクトル系列と呼んでいる。

【0026】次にサブステップSS12では、変換対象話者のコードベクトル系列を生成する。この処理は、選んだ変換対象話者の音声データに対してサブステップSS10、SS11と同様の処理を行って変換対象話者に対するコードベクトル系列を生成する。この生成の後、サブステップSS13に進む。

【0027】サブステップSS13では、マッピングコードブックを生成する。マッピングコードブックは、これま

でに生成した入力話者のコードベクトル系列と変換対象話のスペクトルの特徴を対応付けて集計し、この集計したスペクトルの特徴に重み付け平均することにより、マッピングコードベクトルを生成しこれらマッピングコードベクトルから生成している。このマッピングコードブックの生成の後、サブステップSS14に進む。

【0028】サブステップSS14では、スペクトル変換規則の作成を行う。スペクトル変換規則とは、音声の個人差に関係する特徴の一つであるフォルマント周波数を変換する規則をいう。この規則作成のため、各コードベクトルと各マッピングコードベクトルとにそれぞれにフォルマント分析を施す。これにより、各話者のベクトルに対する第1～第4のフォルマント周波数 $\omega_{c,v1} \sim \omega_{c,v4}$ 、 $\omega_{a,c,v1} \sim \omega_{a,c,v4}$ が4つずつ求められる。これらのフォルマント周波数の各対応関係がスペクトル変換規則として記録される。

【0029】次にサブステップSS15では、入力話者と変換対象話者間のスペクトル変換規則に基づいた変換処理を既に入力されている音声データに施す。このとき、音声データは、既にサブステップSS10でLPC分析を済ませてベクトル量子化されているので、このサブステップでのスペクトル分析には、音声データを用いて高速フーリエ変換（FFT: Fast Fourier Transform）分析だけを行って得られたFFTパラメータをコードベクトルに対応するスペクトル変換規則に従って変換する。この変換により、FFTパラメータ（スペクトル）のフォルマント周波数が変換される。

【0030】次にサブステップSS16では、音声データの作成を行う。すなわち、サブステップSS15で変換されたフォルマント周波数に対応して得られるFFTパラメータから逆変換により音声信号を合成し、この合成された音声声を音声データにして記録する。この結果、合成された音声は、既に記録済みの入力話者と変換対象話者の声質を有する音声となり、入力話者と変換対象話者の中間的な音声となる。

【0031】このような手順を経てリターンに移行してこのサブルーチンSUB1を終了する。これにより、オリジナルな入力音声複数ある場合、2つの音声データを選んでこの処理を繰り返すと、入力された音声に対して適応的な変換を行って高品質な音声声を新たに擬似的に作って記憶させることができる。

【0032】ここで、この新たに生成される音声データの総数を収録済みの音声データとし、この音声データ数が m 個存在するとき、 m 個の中から2つずつ音声データを選んでこれら2つの音声データから1つの新データを生成することは、式(1)の組合せ

【0033】

【数1】

$${}_m C_2 = m! / \{(m-2)! \cdot 2!\} = m! / 2(m-2)! \quad \dots (1)$$

で表される。したがって、音声認識に使用可能な有用な

10

20

30

40

50

音声データ総数は、オリジナルの音声データも含めて、式(2)

【0034】

【数2】

$$m + C_2 = m + m! / \{2 \cdot (m-2)!\} = m(m+1)/2 \quad \dots (2)$$

で表される。ここで、具体例としてオリジナルの音声データ数を20とすると、式(1)あるいは式(2)から音声認識に使用可能な有用な音声データ総数は、210個にデータが大幅に増加することがわかる。

【0035】この方法によれば、オリジナルな音声データ数が少なくても大幅に新たな音声データを短時間で生成することができるようになるので、辞書の作成に要する経済的な負担と時間的な節約という問題を同時に解決することができる。したがって、この方法を適用した辞書を搭載した、たとえば音声認識装置のコストを低減させることができる。

【0036】なお、本発明の実施例では、サブルーチンSUB1でLPC分析を行ってコードブックを作成する手順を説明したが、この手順に限定されるものでなく、後段で処理されているFFT分析もたとえば、サブステップSS10で行ってコードブック内に納めるようにしてもよい。これにより、音声データにおいて煩雑な入出力処理の回数を減らすことができる。

【0037】

【発明の効果】このように本発明の音声認識の辞書作成方法によれば、サンプリングして得られた音声データを用いて音質変換工程で新たな音質を有する音声データを生成し、これらの音声データを基に音声認識学習工程で音声認識を学習させた音声データを保存して辞書を作成して、ある話者からサンプリングされた音声データを用

い新たな話者に対応する音声データが生成され、サンプリングに要する人数を抑え、これによる時間の短縮も図れるので、この辞書作成の経費を大幅に削減させることができる。これにより、この方法を適用した辞書を搭載した、たとえば音声認識装置のコストを低減させることができる。

【図面の簡単な説明】

【図1】本発明に係る音声認識の辞書作成を行うための概略的な辞書作成システムの構成を示すブロック図である。

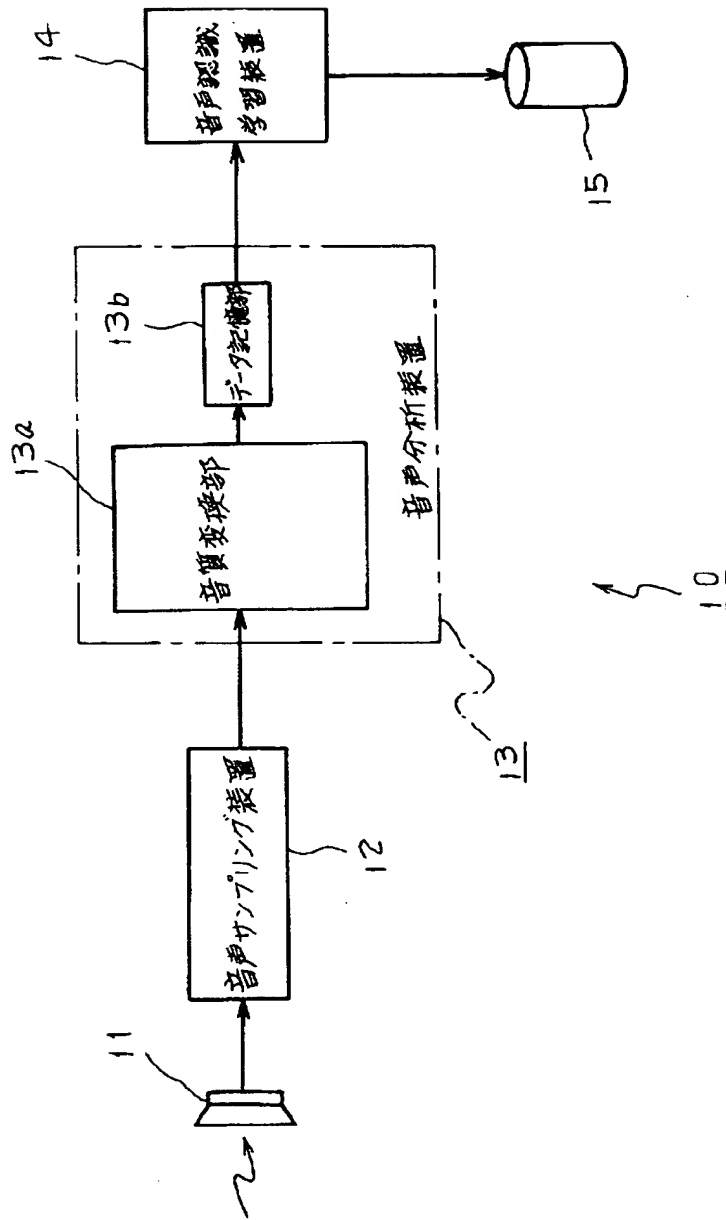
【図2】本発明に係る音声認識の辞書作成方法の一実施例における動作を説明するメインフローチャートである。

【図3】図2に示したサブルーチン1の動作手順を説明するフローチャートである。

【符号の説明】

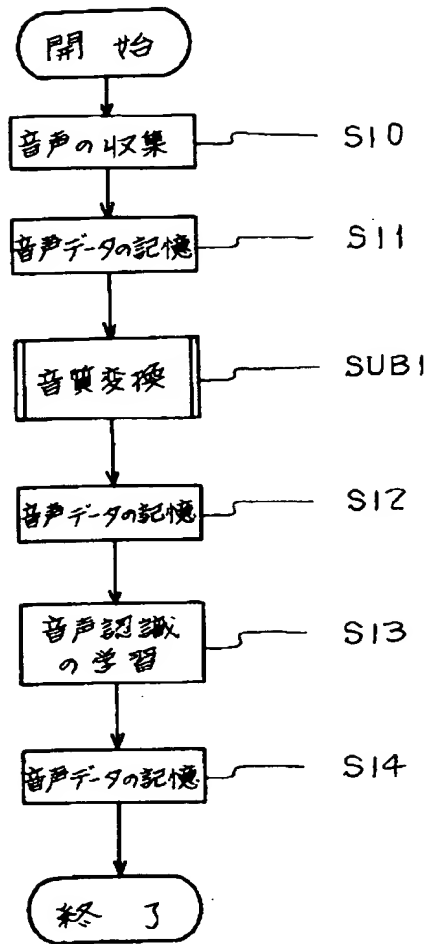
- 10 辞書作成システム
- 11 マイクロフォン
- 12 音声サンプリング装置
- 13 音声分析装置
- 14 音声認識学習装置
- 15 記憶装置（辞書）
- SUB1 サブルーチン1
- SS10 入力話者のコードブックの生成
- SS11 入力話者のコードベクトル系列の生成
- SS12 変換対象話者のコードベクトル系列の生成
- SS13 マッピングコードブックの生成
- SS14 スペクトル変換規則の作成
- SS15 スペクトル変換
- SS16 音声データの作成

【 図 1 】



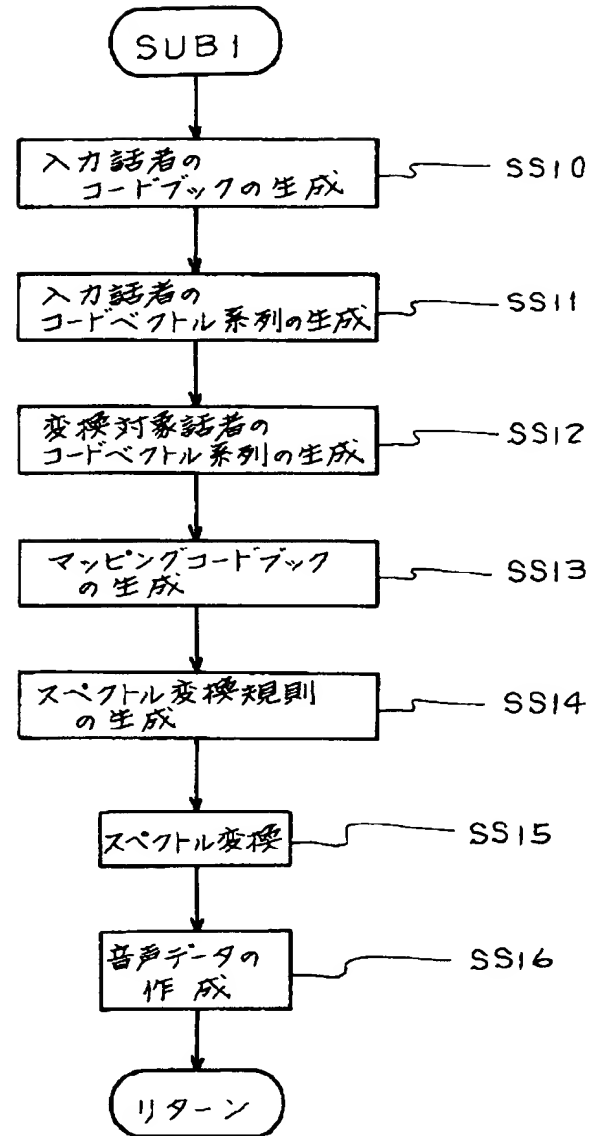
音声認識装置を作成するシステム構成の一例

【図 2】



音声認識の辞書作成方法における
メインフローチャート

【図 3】



サブルーチンのフローチャート